

## Selection on performance and tracking

Citation for published version (APA):

Korthals, R., & Dronkers, J. (2016). Selection on performance and tracking. *Applied Economics*, 48(30), 2836-2851. <https://doi.org/10.1080/00036846.2015.1130789>

**Document status and date:**

Published: 01/06/2016

**DOI:**

[10.1080/00036846.2015.1130789](https://doi.org/10.1080/00036846.2015.1130789)

**Document Version:**

Publisher's PDF, also known as Version of record

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

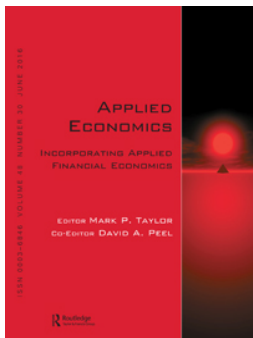
[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.



## Selection on performance and tracking

R. A. Korthals & J. Dronkers

To cite this article: R. A. Korthals & J. Dronkers (2016) Selection on performance and tracking, Applied Economics, 48:30, 2836-2851, DOI: [10.1080/00036846.2015.1130789](https://doi.org/10.1080/00036846.2015.1130789)

To link to this article: <http://dx.doi.org/10.1080/00036846.2015.1130789>



© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 11 Jan 2016.



[Submit your article to this journal](#)



Article views: 128



[View related articles](#)



[View Crossmark data](#)



Citing articles: 2 [View citing articles](#)

## Selection on performance and tracking

R. A. Korthals<sup>a,b</sup> and J. Dronkers<sup>a,b</sup>

<sup>a</sup>Maastricht University, Maastricht, The Netherlands; <sup>b</sup>Research Centre for Education and the Labour Market, Maastricht, The Netherlands

### ABSTRACT

Tracking is widely used in secondary schools around the world. Some countries put more emphasis on the use of performance to place students into tracks (e.g. the Netherlands), while in other countries parents have more influence on the track their child will go to (e.g. Germany). This article examines whether selection into tracks based on performance has an effect on the relation between tracking and student performance and educational opportunities. Using data from the Programme for International Student Assessment for around 185 000 students in 31 countries, different estimation models are compared. The results indicate that a highly differentiated system is best for performance when schools always consider prior performance when deciding on student acceptance. In systems with a few tracks, there is no such impact. Equality of opportunity is best provided for in a system with many tracks when schools always consider prior performance.

### KEYWORDS

Tracking; inequality; student performance; PISA

### JEL CLASSIFICATION

I20; I21; I24

## 1. Introduction

Tracking students in secondary school into distinct educational programmes of different ability levels influences students in a number of ways: In tracked systems, students are separated into tracks with different peers, different curricula and different teachers, while in non-tracked systems in theory all students share their learning environment. Previous research does not show consistent positive or negative effects of tracking on student performance (e.g. Hanushek and Woessmann 2006; Ariga and Brunello 2007; Elk, van Der Steeg, and Webbink 2011). While regarding the inequality of outcomes between students with different parental background (PB; which captures among other things parental education and income), it is often found that tracking reinforces inequality (e.g. Ammermueller 2005; Schuetz, Ursprung, and Woessmann 2008). Although also on inequality there is no consensus (e.g. Walldinger 2006; Brunello and Checchi 2007).

However, a possible explanation for the mixed results found in the literature is that none of the previous papers has looked at how tracking is implemented. For instance, it could be that tracking has different effects on student performance and inequality depending on how students are selected

into tracks. Both the possible unwanted effect of an increased effect of PB, and the possible positive or negative effect on student learning are dependent on how track placement is done: If track placement is done purely on ability levels then the effect of PB is reduced since parents cannot influence track choice directly, while tracks are homogenous in the ability composition which could increase learning. If track placement is not done based on ability, but for instance on PB, then increased learning might not happen since the tracks are not homogenous in ability and the effect of PB naturally increases.

The aim of this article is to investigate whether using prior performance to select students into tracks has an effect on the relation between tracking and student performance and between tracking and educational opportunities. Tracking does not vary within education systems. However, in most countries, school principals are often free in how they select students into tracks. We use data on whether school principals consider prior performance in accepting the students to the school from the Programme for International Student Assessment (PISA) 2009 for 31 OECD countries to investigate our question. We find that students who attend schools whose principals consider prior performance in a highly differentiated

**CONTACT** R. A. Korthals  [r.korthals@maastrichtuniversity.nl](mailto:r.korthals@maastrichtuniversity.nl).

 Supplemental data for this article can be accessed [here](#).

© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

system have higher test scores and a lower impact of PB than those in a comprehensive system. When comparing different estimation models we find it unlikely that our results are driven by the selection which causes better able students to go to schools whose principals consider prior performance in accepting the students to the school.

Some countries have national policies regarding tracking (e.g. Germany, the Netherlands), while others let schools decide whether and how to implement informal tracking (e.g. United States, Sweden). The manner in which students are placed into tracks also differs widely across countries. For instance, in the Netherlands, elementary school students all take an obligatory exit test and combined with the obligatory recommendation of the elementary school teacher on the most suited track, the secondary school accepts students to specific tracks mainly based on that test and that recommendation. In Germany, however, in most northern states, parents have the right to persuade schools to accept their child into the highest tracks, while there is no exit test and the teacher's recommendation is only optional (Dollmann 2011). Higher educated parents are often more willing to ensure their child goes to the higher track, which introduces an advantage for their children. Of these two countries, only the Netherlands places students into tracks based on performance (a proxy of ability combining ability with motivation), while it is to be expected that in Germany a strong effect of PB on track placement, and consequently on student performance, is to be found and this is indeed the case (Dustmann 2004). These two examples show that in the relation of tracking and student performance or educational opportunities, the method of selecting students into tracks could influence outcomes.

To take into account that school principals that consider prior performance possibly do so to be able to accept only the best students to their school, we compare the results of different models which look at between-school between-countries variation, between-school variation and between-countries variation.<sup>1</sup> In the first model, we control for the

best available internationally comparable track level of the individual students. In the second model, we use only the between-schools variation by using country fixed effects, which alleviates possible country heterogeneity, for instance but not exclusively, on track placement policies. And third, by using the national percentage of schools which consider prior performance, we look at between-countries variation only to try to eliminate the bias due to selective student acceptance. That tracking has a positive effect on student performance when the method of track placement is taken into account is seen in all three the models. The coefficients are the smallest in the within country model, while they are the largest, although insignificant, in the between-countries model. If the relation was purely driven by a selection bias created by school principals, this would not be expected since then most of the variation would be between schools in a given country.

The findings indicate that it is important to consider how school principals select students into tracks when looking at the effects of tracking on performance or educational opportunities; or more general, that school characteristics need to be taken into account when analysing education systems. Perhaps this insight can also explain why the literature finds mixed effects of tracking on both student performance and educational opportunities.

The structure of this article is as follows: [Section II](#) discusses theoretical insights in tracking and how schools can affect tracking. [Section III](#) describes the data used in the analysis, while [Section IV](#) lays out the empirical strategy. [Section V](#) presents the results. Finally, the last section concludes the article.

## II. Tracking

Tracking is widely used in secondary school systems around the world. The most important differences between countries in the implementation of tracking are on the number of tracks available to 15 year old students (most frequently ranging from two to five) and on the age of selection into tracks (from 10 to 16). In the countries that have tracking, tracks are

<sup>1</sup>Another reason why schools that consider prior performance might not use this for track placement is that the entry of students into the secondary school might not coincide with the start of tracking. However, in most countries the start of secondary and tracking coincide. Only 7 of the 21 tracking countries do not start tracking at the start of secondary school. Russia's secondary school starts at age 10, but tracking starts at 14.5 years. Students in Luxembourg start secondary school at age 12, but tracking starts at 13. In Lithuania, it is 11 versus 15, in Italy 11 versus 14, in Israel, Ireland and Greece 12 versus 15. However, even for schools in these 7 countries, the earlier obtained prior performance may aid schools in track placement.

institutionalized in different school types and often located in different buildings and administrative units, while in countries without tracking non-institutionalized tracking can occur within schools, either by ability grouping (different classes within schools) or seating (different curricula within classes).<sup>2</sup>

Tracking is a form of imposing peer homogeneity on students, while it also offers students a more targeted curriculum. Each track is aimed to consist of a more or less homogeneous student population, depending on the number of tracks available. However, the effect of imposing peer homogeneity is not theoretically straightforward. First, by removing the better-performing students from the lower tracks, the mean performance of the lower tracks decreases and the resulting lower level of peer performance can harm the performance of the lower ability students. In contrast, the performance of the high-ability students, who are now surrounded by more high-ability peers, improves with positive spillovers. If peer effects work through mean performance, as described above, we would expect to find no country effect of tracking since the positive and negative peer effects on performance cancel each other out. On the other hand, when peer effects are non-linear, tracking can have nationwide effects. The theoretical models of non-linear peer effects support either positive or negative effects of tracking (see Sacerdote 2011). For instance, when especially high-ability students benefit from high-ability peers, tracking has a positive overall effect; when especially low-ability students benefit from high-ability peers, tracking has a negative effect.

Second, peer homogeneity in tracks can be good for both high- and low-ability students when teachers target their teaching to the average performance of the class. In highly differentiated systems (i.e. systems with a large number of tracks), the top and bottom pupils are closer to the average performance level in the class and can thus benefit from peer homogeneity when it allows them to learn more from the teacher.

In addition to imposing peer homogeneity on students, tracking exposes students to specialized curricula, which means that students in different tracks are taught at different levels of difficulty. As

long as the specialized curricula are optimally designed for the average characteristics of the students in the track, they should increase performance.

Overall, we would expect a positive effect of tracking due to improved teaching strategies and adjusted curricula, while the effect of tracking due to more homogeneous peer groups is theoretically uncertain. However, peer effects in the classroom are found to be very context specific, and, if they exist, the size of peer effects is modest (Sacerdote 2014) and thus we disregard them from our expectations. In sum, we therefore expect a positive effect of tracking due to more homogeneous classes. Disentangling the different effects from peers, adjusted curricula and adjusted teacher strategies is not possible in our setup. We will therefore look at all three effects combined.

The arguments in the paragraphs above assume that track placement is based on ability, for instance proxied by prior performance. However, as already mentioned, track placement is not always based on performance. Dustmann (2004) shows for Germany that PB is a strong predictor of track choice and that there is strong intergenerational immobility in track choice. When parents are free to send their child to any of the available tracks, they may choose the track they attended and/or the track they are familiar with. Schools could also select students based on artistic performance or on other aspects as religion or residential area. When this happens, tracks are no longer homogeneous in performance.

We expect that schools that use an objective measure to place students into tracks have greater performance homogeneity in tracks and this induces the expected positive effects of tracking as described above. Schools selecting students into tracks without an awareness of their observed abilities or basing their selection on non-academic criteria can severely limit the expected positive effects of tracking. Schools that have information on prior performance when they place students into tracks, may be better able to ensure homogeneous classes than schools that do not have the same.

We also expect that when schools use a performance measure to select students into track, the influence of parents is lower. Naturally, we do not

<sup>2</sup>In countries with non-institutionalized forms of tracking, placement in the higher performing group can also affect student performance and the effect of PB on performance. However, this non-institutionalized tracking lies outside the scope of this article.

assume that when schools consider prior performance, parents have no influence on school choice or performance. Parents always influence a child's ability, directly through genes and/or indirectly through the environment they create for their children. However, we assume that as long as observed ability limit non-ability-related parental influence on track choice, the effect of PB is decreased.

### III. Data

The student- and school-level data used in this article are from the 2009 wave of the PISA, executed by the OECD. These data include internationally comparable test scores in reading, mathematics and science and information on students and schools. The country-level data are from the OECD and The World Bank.

The first wave of PISA was presented in 2000 and, since then, every 3 years a representative sample of students from all participating countries is subjected to tests on reading, mathematics and science. The test results are standardized to a mean of 500 and a standard deviation of 100 on the PISA reading test in 2000 for the OECD countries.<sup>3</sup> In addition to the tests, the students and school principals are surveyed. A total of 75 countries participated in PISA 2009. Since these countries are diverse in their economic development, we use a selection of comparable Western countries to limit country heterogeneity. All 31 countries in this analysis have a Gross Domestic Product (GDP) per capita above the minimum of the OECD and available data on national tracking policies.<sup>4</sup> These limitations on the sample are imposed to ensure that no country differences drive the results, although also country fixed effects models are used to further take country difference into account.

A representative sample from each participating country is obtained by the OECD in two stages: First, schools are selected and, then, students of the

target age are selected within these schools. The target age is set to a range of 15 years and 3 months to 16 years and 2 months (OECD 2010). Since not all selected schools and students were willing to participate and some schools and students were over-sampled to obtain extra information on these groups, the OECD provides weights to ensure sample representation. The student sample in this analysis consists of all native students in (pre-) vocational or general education who were in schools where more than five students participated in PISA 2009.<sup>5</sup> This amounts to 187 768 students in 7489 schools in 31 countries.

### *Tracking and selection by schools*

We define tracking as the separation of students into tracks that differ in academic orientation and curricula. The extent of tracking is measured using the 'number of school types or distinct educational programmes available to 15-year-olds' (OECD, Table 5.2, 2007), as shown in the first column of Table 1. This measure of tracking is different from those used in some other papers. For instance, Fuchs and Woessmann (2007) and Elk, van Der Steeg, and Webbink (2011) use the age at which a student is first selected into a track as a measure of tracking, while Hanushek and Woessmann (2006) and Schuetz, Ursprung, and Woessmann (2008) divide countries into early versus late trackers. However, this measure of tracking is more consistent with our theoretical expectations from Section II. As discussed in Section II, we expect that the greatest effect from tracking will come from this imposed homogeneity. A disadvantage of using the number of tracks as our tracking measure, however, is that it does not take into account the amount of time students spend in the tracks. Section VI addresses this issue.

Besides a measure on class homogeneity due to tracking (the number of tracks), our framework in

<sup>3</sup>The OECD provides five plausible values estimated using item response theory for the test scores since students do not receive all questions. Here only one plausible value is used.

<sup>4</sup>Countries that are excluded are Australia, Canada, France, Mexico and the United Kingdom. We have excluded France since we have no data on schools, included the needed school data on whether principals consider prior performance. We excluded Turkey since we have no information on the language spoken at home which is an important individual-level background characteristic. The United Kingdom and Australia we excluded since for a large number of observations we do not have data on the proxy for the track the student is in. Canada is excluded since all students are coded as neither vocational nor general education. Mexico is excluded since the values on many variables are outliers: The mean and/or maximum values of the parental background variable and a number of school variables are at least more than half a standard deviation from the mean and/or maximum values of the other countries.

<sup>5</sup>We include only native students since the literature shows that native and migrant students respond differently to system characteristics (e.g. Dronkers, Van Der Velden, and Dunne 2012).



**Table 1.** Tracking and selection by schools.

Country	No. of school tracks available to 15-year-olds	Percentage of students in schools that consider prior performance for student acceptance		
		<i>Never</i>	<i>Sometimes</i>	<i>Always</i>
Argentina	3	47.4	31.3	21.3
Austria	4	21.3	18.2	60.5
Belgium	4	38.3	45.1	16.6
Chile	2	17.4	43.1	39.5
Croatia	3	0.0	6.5	93.5
Czech Republic	5	23.7	27.0	49.3
Denmark	1	50.6	44.6	4.8
Estonia	1	13.5	56.6	30.0
Finland	1	69.2	27.0	3.8
Germany	4	10.3	15.7	74.1
Greece	2	55.0	38.8	6.2
Hungary	3	5.4	4.8	89.8
Iceland	1	69.8	27.4	2.8
Ireland	4	43.7	36.8	19.6
Israel	2	11.1	35.1	53.8
Italy	3	28.6	30.1	41.3
Latvia	3	44.6	30.4	25.0
Lithuania	3	38.9	48.1	13.0
Luxembourg	4	0.5	57.5	42.0
Netherlands	4	0.6	11.7	87.7
New Zealand	1	35.3	39.6	25.2
Norway	1	75.8	18.3	5.8
Poland	1	34.9	49.0	16.1
Portugal	3	71.1	27.7	1.2
Russian Federation	3	36.3	40.8	22.9
Slovak Republic	5	18.3	18.7	63.1
Slovenia	3	23.0	46.6	30.4
Spain	1	79.6	18.0	2.5
Sweden	1	78.6	18.4	3.1
Switzerland	4	22.6	13.1	64.3
United States	1	45.8	26.4	27.8
One track	1 (61 535 students)	55.3 (37 794)	32.5 (17 731)	12.2 (6010)
Two tracks	2 (14 336 students)	27.8 (3748)	39.0 (5634)	33.2 (4954)
Three tracks	3 (66 376 students)	32.8 (21 322)	29.6 (19 943)	37.6 (25 111)
Four tracks	4 (35 217 students)	19.6 (8326)	28.3 (9184)	52.1 (17 707)
Five tracks	5 (10 304 students)	21.0 (1902)	22.8 (2375)	56.2 (6027)

Source: OECD (2007) (first column) and PISA 2009 (second till fourth column). Whether schools consider prior performance is obtained from a questionnaire filled out by the school principal (see text for more details).

Section II requires a school-level variable depicting the selection mechanism with which students are selected into tracks. PISA 2009 contains a proxy for such a variable, namely an index based on a PISA 2009 question to school principals on how often consideration was given to a student's record of academic performance (including placements tests) and to feeder school recommendations in admitting the student to the school. Schools are divided into three categories: (1) schools where neither of the two factors is considered, (2) schools where at least one of these factors is sometimes used to decide acceptance and (3) schools where at least one of the two factors is always considered. In this article, whether schools consider prior performance on acceptance to the school is used synonymous to whether schools have performance criteria for track placement of students. This assumes that secondary schools that have prior performance information use this to decide on the track placement of students.

Table 1 gives an overview of the percentage of schools per country that consider prior performance. Countries differ substantially on the percentage of schools that consider prior performance, from 79.6% of students in schools in Spain that never consider them to 93.5% that always consider them in Croatia. The type of school that considers prior performance also differs across countries: for instance, in the Czech Republic and Hungary, schools that do are often upper secondary schools, while in Austria and Poland, it are mostly schools where students with a high PB attend. In general, village schools or schools without neighbouring schools are less likely to consider prior performance; schools that service more girls, vocational students or students in upper secondary school, higher PB schools and schools which have more teacher shortages are more likely to consider prior performance.

As can be seen from Table 1, in almost every country there are schools in all three categories,

that is, schools that never, sometimes and always consider prior performance. Thus, even in systems with a high number of tracks, some schools do not use prior academic performance or teacher recommendations to decide school admittance. In the seven countries with four tracks, only 55% of the students go to schools that always consider prior performance; in the two countries with five tracks, 18% of students go to schools that never consider prior performance. Maybe more surprising, even in a comprehensive system, some schools consider prior performance when accepting students: In the 10 countries with only one track, 45% of students went to schools that consider prior academic performance. A reason could be that also in countries without tracking, non-institutionalized forms of tracking (ability grouping or seating) exist, which could induce schools to select students based on prior performance. The mechanisms in those countries could work in similar ways as described here. For instance, Lucas (1999) has shown that various methods of placement in non-institutionalized tracks in schools in the US can produce variation in the strength of the effect of prior performance and PB on student performance.

### Control variables

The control variables used represent a standard set of variables used in the literature. All student variables are collected through student surveys. This study controls for gender, age, PB,<sup>6</sup> a dummy for (pre-) vocational education as opposed to general education and a dummy for upper secondary school as opposed to lower secondary education. PB is captured by a widely used index composed by the OECD that describes the student's economic, social and cultural status. The division between upper and lower secondary schools is based on the International Standard Classification of Education level, which provides internationally comparable standards for comparing education levels.

The school-level variables are collected through a survey completed by the school principals. School

composition is captured by the school average and standard deviation of the PB of all the students per school, and by the percentage of them who speak a language other than the test language at home.<sup>7</sup> School inputs are captured by the student-teacher ratio, an index of possible teacher shortages, dummy variables indicating whether the school is hindered by a shortage in instruction material and an index indicating whether the school is responsible for the curriculum and assessment. Other school characteristics, all obtained from the school principal survey, indicate whether school achievement is tracked by an educational authority; whether the school is a public, private government-dependent, or private government-independent school; whether the school has to compete with none, one, or two or more schools for students; the school location; school size; and whether the school uses ability grouping. We also control for the GDP per capita PPP (constant 2005 international dollars), which is for 2008 from the World Bank (2012). Table 2 provides descriptive statistics for all variables.

### IV. Empirical strategy

The aim of this article is to investigate whether using performance to select students into tracks changes the relation between tracking and student performance and educational opportunities. To answer this question we make use of three models: a between-schools between-countries model depicted in Equation (1), a within-country model depicted in Equation (2) and a between-countries model depicted in Equation (3).

$$\begin{aligned} Test_{isc} = & \beta_0 + Student_{isc}\beta_1 + School_{sc}\beta_2 \\ & + EntrReq_{sc}\beta_3 + No. of Tracks_c\beta_4 \\ & + EntrReq_{sc} \times No. of Tracks'_c\beta_5 \\ & + GDPpc_c\beta_6 + u_c + u_{sc} + \epsilon_{isc} \end{aligned} \quad (1)$$

$$\begin{aligned} Test_{isc} = & \delta_0 + Student_{isc}\delta_1 + School_{sc}\delta_2 \\ & + EntrReq_{sc}\delta_3 + C_c\delta_4 + EntrReq_{sc} \\ & \times No. of Tracks'_c\delta_5 + w_{sc} + \vartheta_{isc} \end{aligned} \quad (2)$$

<sup>6</sup>PB is an important control variable since it is very well established that parents have a large influence on student performance. Among our 31 countries, the correlation between PB and the reading score is between 0.23 and 0.49, with the lowest score in Iceland and the highest in Hungary. Countries without tracking at age 15 have a correlation of 0.30, while countries with tracking have a correlation of 0.37.

<sup>7</sup>To calculate the school PB composition, both native and immigrant students are used.



**Table 2.** Descriptive statistics.

Variable	Mean	SD	Minimum	Maximum	Type	% missing
PISA reading score	493.11	93.52	1.97	817.42	cont	0.00
PISA mathematics score	495.33	94.61	42.85	901.86	cont	0.00
PISA science score	502.88	94.98	10.95	863.24	cont	0.00
Parental background	0.07	0.94	-5.34	3.41	cont	0.38
Gender	0.50	0.50	0.00	1.00	dum	0.00
Age	15.77	0.29	15.25	16.33	cont	0.00
Student in (pre-) vocational education	0.16	0.37	0.00	1.00	dum	0.01
Student in upper secondary school	0.47	0.50	0.00	1.00	dum	0.01
Mean school PB	0.04	0.56	-2.55	1.61	cont	0.00
SD of school PB	0.77	0.16	0.17	1.58	cont	0.00
Percentage in class for whom the first language is not the test language	4.90	1.25	1.00	6.00	cat	2.70
Student-teacher ratio	12.72	6.45	0.39	264.33	cont	8.24
Teacher shortage	-0.14	0.88	-1.02	3.34	cont	2.08
Hinder by a shortage instruct material	1.82	0.83	1.00	4.00	cat	2.30
School responsible for curriculum and assessment	-0.14	0.94	-1.37	1.36	cont	1.56
Index of achievement tracked by authority	0.34	0.47	0.00	1.00	dum	3.36
School type	1.18	0.44	1.00	3.00	cat	3.57
School competition	1.67	0.85	1.00	3.00	cat	2.07
School location	2.88	1.09	1.00	5.00	cat	1.60
School size	659.48	491.29	10.00	6850.00	cont	3.27
Ability grouping	1.78	0.67	1.00	3.00	cat	0.36
School considers prior performance	1.98	0.83	1.00	3.00	cat	2.77
No. of tracks	1.61	1.31	0.00	4.00	cont	0.00
GDP per capita, 2008	\$28 856.03	\$12 410.14	\$13 275.68	\$73 349.64	cont	0.00

Source: PISA (2009), OECD (2007: Number of tracks available to 15-year old students) and World Bank (2012: GDP per capita 2008).

Variables can be continuous (cont), dummy variables (dum) or categorical variables (cat). The mean and standard deviation of the school PB is constructed from the imputed PB data. The table presents data that are not demeaned.

$$\begin{aligned}
 Test_{isc} = & \theta_0 + Student_{isc}\theta_1 + School_{sc}\theta_2 \\
 & + Nat\%EntrReq_c\theta_3 \\
 & + No. of Tracks_c\theta_4 \\
 & + Nat\% EntrReq_c \\
 & \times No. of Tracks'_c\theta_5 + GDPpc_c\theta_6 \\
 & + z_c + z_{sc} + \mu_{isc}
 \end{aligned} \quad (3)$$

In these equations,  $Test_{isc}$  is the individual PISA test score in reading, mathematics or science of student  $i$  in school  $s$  in country  $c$ .  $Student_{isc}$  is a matrix of student variables,  $School_{sc}$  is a matrix of school variables, while  $EntrReq_{sc}$  is a matrix containing the dummies on whether schools consider prior performance.  $No. of Tracks_c$  is a vector containing the number of tracks available to students in each country.  $GDPpc_c$  is a vector containing GDP per capita. Our main coefficients of interest of Equation (1) are  $\beta_3$ ,  $\beta_4$  and  $\beta_5$  which capture the effect of the number of tracks combined with whether the school principal considered prior performance.<sup>8</sup> As compared to the main model in Equation (1), model (2) adds country fixed effects,  $C_c$ , while in model (3) we use the

national percentage of school whose principals sometimes or always consider prior performance on accepting the student to the school,  $Nat\% EntrReq_c$ , as opposed to the school-level variable. We use random effect models, which are estimated using maximum likelihood, to take into account error terms for countries, schools and individuals: Separate error terms are therefore included for countries, schools and individuals, since students are nested within schools within countries. Weighting is used to ensure representative samples.<sup>9</sup> The control variables are like discussed in Section IV.

The reason we analyse three models is that we are concerned with multiple selection issues. First, we add to our standard model (Equation (1)) also a country fixed effects model (Equation (2)) to exclude the possibility of country heterogeneity biasing the results. By using a country fixed effect model, we only compare outcomes of students in schools in which the principals does or does not consider prior performance, given the country and given the number of tracks which are available in that country. The interaction between the number of tracks and whether principals

<sup>8</sup>Not all students are tracked at the same time: the age of selection runs from age 10 to age 15. Therefore, the effect of the number of tracks and whether school principals considered prior performance on students' performance may have occurred at the time when the PISA tests were administered (age 15–16), but it could also have occurred earlier in time and we only capture it at age 15 when the students do the test.

<sup>9</sup>The weights are adjusted such that each country has an equal weight in the estimation. All statistics and estimations are executed using weights, unless otherwise specified.

consider prior performance gives us the estimates we are interested in. The main tracking effect, which drops out, will be captured in the country fixed effects. Second, we use also between-country analyses (Equation (3)) to exclude the alternative explanation of why school principals would consider prior performance, namely to select the best students and thus the best performances for their school ('cherry picking').

Even given our alternative models, it could still be that all countries with a lot of tracks and which have schools that select students are different from all countries that do not. Our country-fixed effects models would not be able to control for this. Furthermore, although we control for a large number of school characteristics and also check for sorting into schools using our between-country analyses, unobservables which are related to student performance, tracking and whether schools select students based on performance and which would cancel out on the country level could potentially exist.

To study the hypotheses on student performance, the main focus lies on the interaction between tracking and whether schools consider prior performance,  $EntrReq_{sc} \times no. of Tracks'_c$ . Our hypothesis is that this interaction will be positive: School principals that consider prior performance in selecting students into tracks are more likely to achieve positive tracking effects due to more homogenous classes.

To look at whether the effect of PB is lower when track placement is done based on prior performance we use Equations (4) and (5). Equation (4) contains an interaction between PB and the number of tracks in a country, and will show whether in countries with more tracks there is a larger effect of PB. Equation (5) then adds the interaction between number of tracks, PB and whether school principals consider prior performance. This triple interaction will show whether the effect of PB is lowered when track placement is done based on ability. The PB of students is included in all models in the vector,  $Student_{isc}$ .

$$\begin{aligned} Test_{isc} = & \gamma_0 + Student_{isc}\gamma_1 + School_{sc}\gamma_2 \\ & + EntrReq_{sc}\gamma_3 + No. of Tracks_c\gamma_4 \\ & + PB_{isc} \times No. of Tracks'_c\gamma_5 \\ & + GDPpc_c\gamma_6 + v_c + v_{sc} + \varepsilon_{isc} \end{aligned} \quad (4)$$

$$\begin{aligned} Test_{isc} = & \rho_0 + Student_{isc}\rho_1 + School_{sc}\rho_2 \\ & + EntrReq_{sc}\rho_3 + No. of Tracks_c\rho_4 \\ & + PB_{isc} \times No. of Tracks'_c\rho_5 + PB_{isc} \\ & \times No. of Tracks'_c \times EntrReq_{sc}\rho_6 \\ & + GDPpc_c\rho_7 + t_c + t_{sc} + \pi_{isc} \end{aligned} \quad (5)$$

Almost all student- and school-level variables have some missing observations (see last column of Table 2). Although most variables have below 3% missing values, deleting all observations with missing variables would lead to a drop in observations from around 185 000 to around 130 000, and it would assume that the missing values are missing at random, which is a questionable assumption. Another reason for not deleting all observations is that it leads to distorted weighting. Therefore, the missing values in the sample are replaced by group averages.<sup>10</sup> To control for possible bias introduced by the method for replacing missing values, imputation dummies and imputation interactions are used in all models.<sup>11</sup>

## V. Results

First, we replicate the standard cross country analysis of the effect of tracking on student performance. Then, we turn to investigating whether using performance to select students into tracks has an effect on the relation between tracking and student performance. We first present the results on the between-countries between-schools models, and subsequently show the between-countries and the within-country results. And third, we focus on whether using performance to select students into tracks has an effect on the relation between tracking and educational opportunities.

### Direct relation between tracking and student performance

The analysis starts by investigating whether the number of tracks has a direct and significant relation with student performance. Since we include a wide variety of school background variables, which may capture part of the tracking effect, we do not expect

<sup>10</sup>The student variables are replaced by the average value of the students in the same school, the school variables are replaced by the country average. Country variables are never missing.

<sup>11</sup>The results are robust to the exclusion of the imputation variable interaction terms, and imputation dummies.

**Table 3.** The effect of tracking on performance.

Dependent variable	(1) Reading	(2) Mathematics	(3) Science
School considers prior performance			
<i>Sometimes</i>	−2.02 (1.95)	−1.37 (1.89)	−1.13 (2.06)
<i>Always</i>	6.34* (3.27)	5.51* (3.07)	7.45** (3.79)
No. of tracks (0–4)	−2.56 (2.55)	1.94 (3.67)	0.75 (3.30)
(Pseudo-)LL	−43 761	−43 739	−44 094
No. of students	187 768	187 768	187 768
No. of countries	31	31	31

The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and the number of tracks in a country, controlling for whether or not schools consider prior performance when selecting students. The superscripts \* and \*\* indicate significance at the 10% and 5% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

a large coefficient for tracking. Table 3 confirms our expectations: The association between the number of tracks and performance is insignificant, while it is negative for reading and positive for mathematics and science. For all three test subjects, the relation between schools that sometimes consider prior performance and student performance is negative, while for schools that always consider prior performance this relation is positive (and significant).

All the control variables, which are excluded from the table, have the expected sign. The results of the full models are available upon request.

### Tracking and performance

To test whether considering prior performance to select students into tracks has an effect on the relation between tracking and student performance, we include interactions between whether schools consider prior performance and the number of tracks, as in Equation (1). The results can be seen in Table 4. More tracks in an education system are positive for student's performance if students attend schools where the principals consider prior performance on accepting the student to the school. For reading there is still a significant negative effect of more tracks (−5.74\*\*), but this is compensated when schools always consider prior performance and there are 4 or 5 tracks to choose from. Since for mathematics and science no significant negative coefficient of tracking exists (−0.88 and −2.89), more tracks are even better for students when they attend schools which always considers prior performance.

**Table 4.** The effect of tracking on performance and the influence of selection on performance.

Dependent variable	(1) Reading	(2) Mathematics	(3) Science
School considers prior performance			
<i>Sometimes</i>	−5.97** (2.54)	−4.95** (2.40)	−5.99** (2.44)
<i>Always</i>	−10.13** (4.11)	−9.54** (4.25)	−10.90** (5.06)
No. of tracks (0–4)	−5.74** (2.70)	−0.88 (3.75)	−2.89 (3.31)
<i>Sometimes</i> × No. of tracks	2.83** (1.15)	2.50** (1.16)	3.45*** (1.08)
<i>Always</i> × No. of tracks	8.64*** (1.82)	7.85*** (1.90)	9.68*** (2.04)
(Pseudo-)LL	−43 759	−43 738	−44 092
No. of students	187 768	187 768	187 768
No. of countries	31	31	31

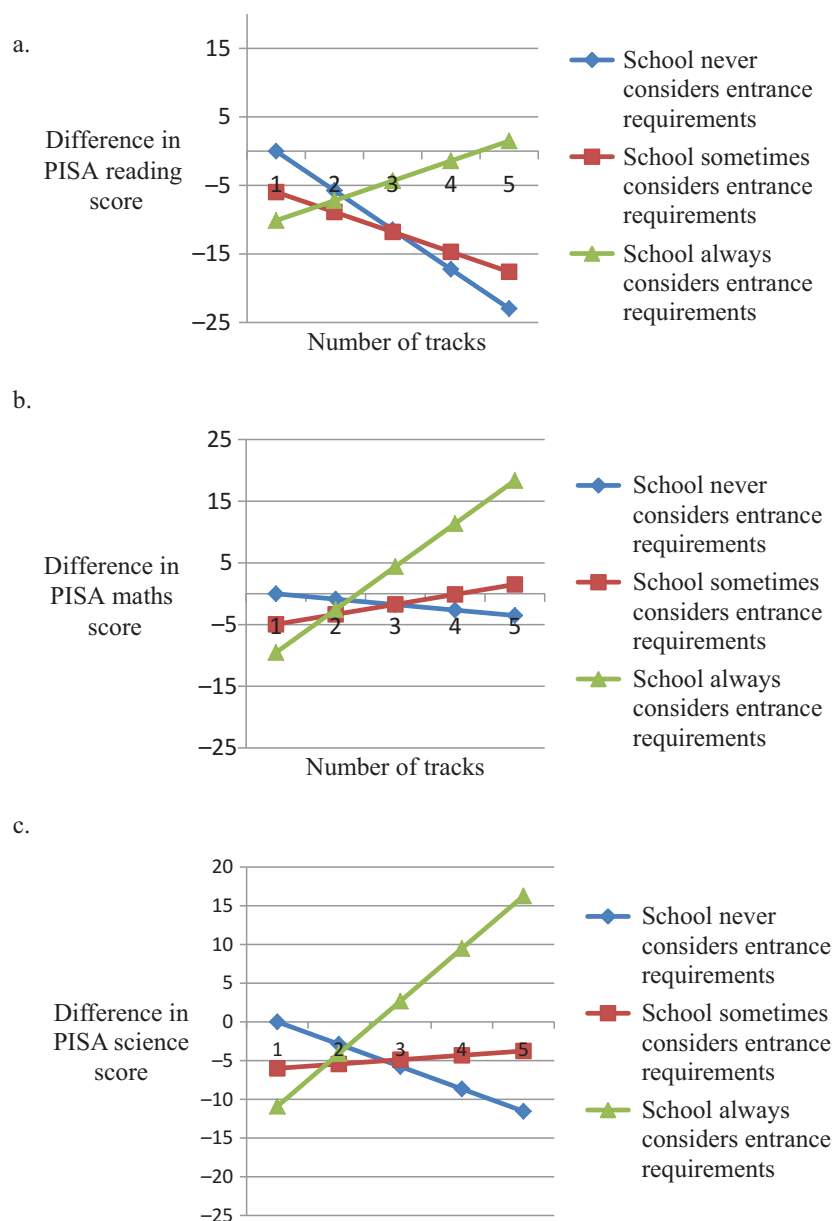
The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and whether or not schools consider prior performance when selecting students and the number of tracks in a country. The superscripts \*\* and \*\*\* indicate significance at the 5% and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

To facilitate the interpretation of the interaction terms, Figure 1 shows the combined coefficients for the three models. Figure 1 shows for each combination of number of tracks and whether schools consider prior performance what the relation between the two and student performance is, relative to students in a system with only one track in schools that never consider prior performance. Looking at the figures, one sees the same trend for all three subjects (reading, mathematics and science): schools in multiple track systems do better when they consider prior performance more often, while schools in comprehensive systems perform better when they do not consider prior performance. When only the significant differences in the graphs are considered, it becomes clear that for two or more tracks whether schools consider prior performance only changes the results when schools always consider prior performance. The coefficients for Never and Sometimes are not significantly different from each other when the number of tracks is two or more.<sup>12</sup>

The models in Table 4 use the number of tracks in a country as a continuous variable ranging from zero to four. Appendix A found in the online Supplemental material contains non-linear models which include the number of tracks in a country as dummy variables. The results are qualitatively similar.

From Table 4 it can be concluded that students in a system with a high number of tracks do better when their school always considers prior performance, while for students in a system with a low number of

<sup>12</sup>Except for science, where the combined coefficients for the Never and Sometimes are significantly different in a system with five tracks.



**Figure 1.** Differences in student performance for students in different education systems that attend schools that do or do not consider prior performance, relative to students in a comprehensive system in schools that never consider prior performance. (a) Reading, (b) Mathematics and (c) Science.

Source: Own calculations using PISA (2009).

tracks (two or three) whether schools consider prior performance do not seem to matter. In a system with four or five tracks, the schools that consider prior performance can place students into the available tracks based on this information, and thus in these schools class homogeneity is higher as compared to schools that do not obtain information on prior performance of their students. The data suggest that to place students into only two or three tracks is not beneficial for student performance regardless of whether schools consider prior performance. A

possible explanation for this is that two or three tracks do not allow for enough differentiation between students with heterogeneous ability. In a system with one track whether schools consider prior performance still matters for student performance: Although students in schools that sometimes or always consider prior performance do not perform differently from each other, schools that never consider prior performance perform (marginally) better.

The results presented above will be biased if the variation in the data is affected by sorting of students

into schools. This may be the case when schools that consider prior performance on accepting students do so to be able to select the most able students and not to allocate students into tracks. Under the assumption of full sorting into schools, it is to be expected that schools that consider prior performance will have students that always perform better. However, this is not the case: In systems with only one track, whether schools consider prior performance does not seem to matter much (it is only marginally better to be in a school that does not consider prior performance than to be in one which does) and in a system with only a few tracks (2 or 3), whether schools consider prior performance do not seem to matter at all. Therefore, schools which consider prior performance when deciding on accepting the student to the school do not perform better by definition.

That the hypotheses following full sorting do not seem to be confirmed by the data does not mean sorting is not a problem in these analyses. Sorting in lesser extent can still exist and could potentially bias the results. To investigate this, we present results using only between-countries variation, as shown for mathematics in Table 5. Appendix B found in the online Supplemental material shows the same comparison for reading and science. The first column of Table 5 replicates column (2) from Table 4 for comparative purposes. The third column of Table 5 replaces the school-level dummies on whether schools consider prior performance by variables depicting the national proportion of students in schools that sometimes or always consider prior performance. This model excludes the possible sorting of students into schools since this micro-phenomenon cannot intervene with the estimation when whether schools consider prior performance is measured at the country level. Schools that always consider prior performance have a negative impact on student performance ( $-134.23^{***}$ ); however, this is compensated in countries with more than 3 tracks due to the positive interaction term with the number of tracks ( $49.89^{***}$ ).<sup>13</sup> For an average country, where

**Table 5.** Comparing different models on student performance (mathematics as dep. var).

	(1) Between countries, between schools	(2) Within country	(3) Between countries
School considers prior performance			
<i>Sometimes</i>	$-4.95^{**}$ (2.40)	$-3.82^*$ (2.03)	$99.29^*$ (54.46)
<i>Always</i>	$-9.54^{**}$ (4.25)	$-2.03$ (3.99)	$-134.23^{***}$ (40.72)
No. of tracks (0–4)	$-0.88$ (3.75)	–	$9.60$ (17.72)
<i>Sometimes</i> × No. of tracks	$2.50^{**}$ (1.16)	$2.28^{**}$ (1.01)	$-58.77$ (37.00)
<i>Always</i> × No. of tracks	$7.85^{***}$ (1.90)	$5.11^{**}$ (2.01)	$49.89^{**}$ (22.7)
Country FE	–	√	–
(Pseudo)LL	$-43\ 738$	$-43\ 695$	$-43\ 733$
No. of students	187 768	187 768	187 768
No. of countries	31	31	31

The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and whether or not schools consider prior performance when selecting students and the number of tracks in a country using three specifications. Column (1) shows the main model as depicted in column (2) of Table 4. Column (2) shows the same models but with country fixed effects included. Column (3) measures the school variables 'school principal consider prior performance' on a national level, and thus depicts the proportion of schools (between 0 and 1) in the country with school which say they always or sometimes consider prior performance. The superscripts \*, \*\* and \*\*\* indicate significance at the 10%, 5% and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

29% of students are in schools that sometimes consider prior performance and where 32% are in schools that always consider prior performance, performance for reading is best if there is one track and performance for mathematics and science is best if there are five tracks. For a country with a high percentage of schools that always consider prior performance, students perform best in a five-track system, regardless of the subject.<sup>14</sup> For countries with high numbers of schools that never or sometimes consider prior performance, students perform best with two tracks, also regardless of the subject.<sup>15</sup> There results seem to indicate that, although sorting into schools could be a problem, it is unlikely that it alone drives our results.

We are also concerned that country heterogeneity could influence our results. For that reason Table 5 also contains a country fixed effects model using mathematics as dependent variable. The second

<sup>13</sup>The variables and their interactions do not show significant results for reading and science, partly due to the reduction in the number of degrees of freedom. However, an *F*-test is performed to see whether the main and interaction effects are jointly significant. They are for mathematics and reading. For science the five variables are not jointly significant, but the three variables relating to schools that always consider prior performance are jointly significant at the 5% level.

<sup>14</sup>In this case, 58% in schools that always consider prior performance (the mean plus one standard deviation).

<sup>15</sup>In this case, 42% of students attend schools that sometimes consider prior performance (the mean plus one standard deviation) and 6% attend schools that always consider prior performance (the mean minus one standard deviation).



column of Table 5 shows the same model but now including country fixed effects to check whether country heterogeneity influences the results. The interactions are a bit smaller (2.28\*\* versus 2.50\*\* and 5.11\*\* versus 7.85\*\*), but qualitatively similar. For all subjects, students perform best in systems with a high number of tracks when schools always consider prior performance. As also pointed out by Schuetz, Ursprung, and Woessmann (2008), a model with country fixed effects provides unbiased results for cross-country analysis, assuming that the existing country heterogeneity does not influence the interaction between whether schools consider prior performance and the number of tracks. Although this model still does not allow for a strict causal interpretation, the assumption required is considerably weaker than the assumption that no unobserved country heterogeneity exists, even with a sample of similar countries.

### Tracking and inequality

The results for whether using performance to select students into tracks has an effect on the relation between tracking and educational opportunities are displayed in Table 6. As expected and consistent with the literature, PB has a positive and substantial relation with test scores. This effect is similar over the three PISA test subjects, reading, mathematics and science, and comprises about a quarter of a standard deviation in the test scores. If we look at the interaction between PB and tracking in columns (1)–(3), it can be seen that tracking mitigates the association with PB, and therefore reduces inequality of opportunity: The interaction of the number of tracks and PB is negative and highly significant. The first three

columns show that in a system with five tracks the association of PB is lowered by 17.3 points in reading, 16.2 in mathematics and 18.2 in science.

Finding that tracking reduces inequality is not fully consistent with the literature, which most often finds that tracking increases inequality or has no effect. To further investigate the drivers of this positive effect of tracking, we show in columns (4)–(6) in Table 6 the results of the same models, but now including also interactions between PB, the number of tracks and whether schools consider prior performance on deciding to accept the student to the school. This allows us to check if for schools that consider prior performance, and are assumed to use this information for track placement, the effect of PB is lower. The last three columns show that the interaction between the number of tracks and PB is no longer significant; however, the triple interactions (number of tracks, PB and whether schools consider prior performance) are. The table indicates that it is primarily the schools that always consider prior performance in systems with multiple tracks that mitigate the relation of PB with performance. Thus it is not tracking itself that diminishes the association of PB and performance but, rather, tracking combined with whether schools consider prior performance. This is consistent with our expectation that when schools consider prior performance, parents have less influence on their child's track choice and subsequent performance.

### VI. Robustness

In this section, we consider other possible distorting factors for our analyses: heterogeneous effects for PB groups, the measure of tracking used and the sample

**Table 6.** The effect of tracking on inequality and the influence of selection on performance.

Dependent variable	(1) Reading	(2) Math	(3) Science	(4) Reading	(5) Math	(6) Science
Parental background	23.68*** (1.86)	24.45*** (2.06)	24.76*** (2.02)	23.44*** (1.91)	24.26*** (2.10)	24.53*** (2.07)
School considers prior performance						
<i>Sometimes</i>	–1.91 (1.94)	–1.35 (1.89)	–1.09 (2.05)	–2.51 (2.01)	–1.98 (2.01)	–1.75 (2.06)
<i>Always</i>	6.72* (3.39)	5.7* (3.07)	7.75* (3.74)	5.91* (3.35)	4.96 (3.11)	6.90* (3.71)
No. of tracks (0–4)	–2.50 (2.62)	2.03 (3.73)	0.84 (3.38)	–2.27 (2.59)	2.23 (3.71)	1.06 (3.37)
Parental background × No. of tracks	–3.45*** (0.99)	–3.23*** (1.11)	–3.64*** (1.09)	–1.93 (1.21)	–1.80 (1.25)	–2.05* (1.22)
Parental background × No. of tracks × <i>Sometimes</i>				–0.90** (0.37)	–1.13** (0.55)	–1.13** (0.52)
Parental background × No. of tracks × <i>Always</i>				–2.49*** (0.72)	–2.18*** (0.72)	–2.50*** (0.67)
(Pseudo-)LL	–43 752	–43 731	–44 085	–43 750	–43 730	–44 083
No. of students	187 768	187 768	187 768	187 768	187 768	187 768
No. of countries	31	31	31	31	31	31

The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and parental background, whether or not schools consider prior performance when selecting students and the number of tracks in a country. The superscripts \*, \*\* and \*\*\* indicate significance at the 10%, 5% and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.



of countries. We report all robustness checks with the PISA mathematics score, unless otherwise stated the results for reading and science are robust to the various checks.<sup>16</sup> The results which are not shown are available on request.

### Sample split on PB

Table 6 indicates that the differences between students of low and high PB are minimized in a highly differentiated system. To investigate whether the effects of tracking and whether schools consider prior performance on accepting students on equality of opportunity are indeed different for students of different socio-economic background, we estimate models for the subsamples of low and high socio-economic background students. Table 7 shows the results. For students with high PB (column (2)) the relation between tracks and performance is negative, irrespective of whether schools consider prior performance. However, for low PB students (column (1)) the number of tracks does not alter educational opportunities. However, the triple interactions (PB, number of tracks and whether schools consider prior performance) show that when schools always consider prior performance, a high number of tracks are beneficial for low PB students since their

disadvantaged background has less of an effect on their performance.

### Timing of tracking

Unlike some other papers on this topic, we use the number of tracks available to students at the age of 15 to characterize a country's tracking regime. Since the PISA test is conducted when the students are between 15 and 16, it is possible that although multiple tracks are available to students at the age of 15, students have not yet been tracked for a substantial amount of time. If this is the case, the association with tracking in late selection countries may be too weak to be picked up. To check for this, we redo the analysis including interactions between whether a country selects early, the number of tracks, and whether school principals consider prior performance. Table 8 shows the results for the mathematics score as dependent variable, using three definitions of early tracking: tracking before the age of 13, before 14 or before 15. Table 8 shows that only for the students in early tracking countries our results hold since only the triple interaction early tracking, the number of tracks and prior performance is always considered as significant. However, if we use the reading- or science score, then also the interaction for number of tracks and prior performance is always considered is significant (not shown). Our results seem, therefore, to be strongest for countries that track early, but are not absent for countries that track later.

**Table 7.** Robustness checks: Sample split on PB (mathematics score as the dependent variable).

Model:	(1) Low PB	(2) High PB
School considers prior performance		
<i>Sometimes</i>	−2.86 (2.12)	−0.70 (1.76)
<i>Always</i>	3.65 (3.16)	7.64** (3.37)
No. of tracks (0–4)	2.84 (4.09)	4.04 (3.72)
Parental background × No. of tracks	−0.50 (1.78)	−3.08** (1.31)
Parental background × No. of tracks × <i>Sometimes</i>	−1.17 (0.74)	−0.36 (1.13)
Parental background × No. of tracks × <i>Always</i>	−2.54** (1.19)	−0.86 (0.93)
(Pseudo) LL	−43 311	−43 918
No. of students	95 406	92 362
No. of countries	31	31

The table presents coefficients from random effects models (standard errors in parenthesis). Column (1) and (2) replicate column (5) of Table 6 but column (1) uses only low PB students, while column (2) uses only high-PB students. The superscript \*\* indicates significance at the 5% level. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

### Excluding countries without tracking

Even though we estimate country fixed effects models alongside our main models in Table 4, there could still be concern about difference among countries with different tracking regimes which could affect the relation between tracking and student performance, and the relation between PB, tracking and student performance. What could especially be of interest is whether countries that do not track at all are very different from the countries that do track. Therefore, we have also estimated our models without the countries that do not track their students.

<sup>16</sup>Most of the changes are for students in schools that always considers prior performance. The main result (schools that consider prior performance for student acceptance mitigates the relation between strong tracking and performance) holds in any specification.

**Table 8.** Robustness checks: Timing of tracking (mathematics score as dependent variable).

	(1) Early < 13	(2) Early < 14	(3) Early < 15
School considers prior performance			
<i>Sometimes</i>	−3.60 (2.33)	−4.56** (2.31)	−4.82** (2.32)
<i>Always</i>	−5.43 (3.81)	−6.65 (4.13)	−8.22** (3.85)
No. of tracks (0–4)	−1.07 (3.71)	−1.11 (3.74)	−1.06 (3.74)
<i>Sometimes</i> × No. of tracks	−0.20 (2.35)	1.56 (2.05)	2.02 (2.31)
<i>Sometimes</i> × No. of tracks × Early	3.62 (2.20)	1.36 (2.17)	0.68 (2.58)
<i>Always</i> × No. of tracks	1.00 (2.26)	2.43 (2.61)	2.50 (2.49)
<i>Always</i> × No. of tracks × Early	7.41*** (1.94)	5.74*** (2.22)	5.62** (2.70)
No. of students	187 768	187 768	187 768
No. of early tracking countries	9	11	15
No. of late tracking countries	22	20	16

The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and whether or not schools consider prior performance when selecting students and tracking in a country. Column (1) uses as definition for early tracking country if the students are tracked before the age of 13, Column (2) uses before the age of 14, while Column (3) uses before the age of 15. The superscripts \*\* and \*\*\* indicate significance at the 5% and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

This leaves us with 21 countries and 126 333 students. The results are shown in Table 9. What we find when we exclude countries which do not track in Table 9 is that the effect sizes of the coefficients

even larger than in Table 4. Only the coefficients from the between-countries model (Column (3)) are very different from before: The coefficients are closer to zero and far from significant. These results are very similar over the three test subjects.

For the models on PB where only the interaction between PB and number of tracks is included, Column (4), the coefficient is half the size as before and is no longer significant. But including the triple interaction, Column (5), the results are very similar to Table 6. Also the PB results do not seem to be very different when non-tracking countries are excluded.

### Excluding other countries

Using dummies for the number of tracks as in Appendix A found in the online Supplemental material reveals that especially countries with only a few tracks perform worse. Looking at the countries with only a few tracks, the three countries with two tracks (Israel, Greece and Chile) are among the worst-performing countries with regard to PISA test scores in reading, mathematics and science. However, excluding these countries does not influence the results much (not shown).

**Table 9.** Robustness checks: Excluding non-tracking countries (mathematics score as dependent variable).

	(1) Between countries, between schools	(2) Within country	(3) Between countries	(4) PB	(5) PB-triple interaction
Parental background	14.87*** (1.64)	14.24*** (0.35)	14.89*** (1.64)	11.20*** (4.02)	9.66** (3.87)
School considers prior performance					
<i>Sometimes</i>	−8.37 (7.01)	−3.99 (4.89)	−52.19 (168.00)	−0.26 (2.75)	−1.58 (3.03)
<i>Always</i>	−19.71** (9.07)	−12.55** (5.40)	−56.09 (85.87)	6.70* (3.47)	5.23 (3.61)
No. of tracks (1–4)	16.32*** (5.20)	–	20.42 (42.75)	21.86*** (4.55)	22.10*** (4.52)
<i>Sometimes</i> × no. of tracks	3.34 (2.40)	2.93 (2.01)	−14.61 (84.57)		
<i>Always</i> × no. of tracks	10.53*** (3.58)	9.81*** (2.10)	7.057 (42.07)		
Parental background × No. of tracks				1.55 (1.64)	3.93** (1.58)
Parental background × No. of tracks × <i>Sometimes</i>					−1.40** (0.65)
Parental background × No. of tracks × <i>Always</i>					−2.94*** (0.58)
Country FE	–	√	–	–	–
No. of students	126 233	126 233	126 233	126 233	126 233
No. of countries	21	21	21	21	21

The table presents coefficients from random effects models (standard errors in parenthesis) on the relation between student performance and parental background, whether or not schools consider prior performance when selecting students and the number of tracks in a country. The superscripts \*, \*\* and \*\*\* indicate significance at the 10%, 5% and 1% levels, respectively. The control variables are as described in the text. The models include imputation dummies and imputation variable interaction terms.

## VII. Conclusion

The variation in tracking in education systems throughout the Western world is quite large: Many countries have no tracking in secondary school (although most have some form of non-institutionalized tracking), while some countries distinguish up to five tracks for students. Also the manner in which tracking is implemented on the school level differs widely. Some countries put more emphasis on the use of performance to place students into the available tracks (e.g. the Netherlands), while in other countries parents have more influence on the track their child will go (e.g. Germany). We argue that the inconsistencies in the empirical results found in the literature could be explained by country differences in track placement. When track placement is done not based on an ability measure but mainly on PB, or, related, residential areas, the theoretical benefits of tracking might not arise. These theoretical benefits rely heavily on the idea that tracking leads to more homogenous ability classes which the teacher is better able to teach and where students benefit from a curriculum tailored to their needs and abilities. When students are placed into tracks based not on prior performance but on PB, classes will be more heterogeneous and students might not be taught a fitting curriculum. Furthermore, when PB is used to decide on track placement, educational opportunities are affected and inequality is likely to grow.

To study whether using performance to select students into tracks has an effect on the relation between tracking and student performance and educational opportunities, we use data of around 185 000 students in 31 comparable countries from PISA 2009. Prior performance can be thought of as an important measure of observed student ability. Therefore, for schools having the information on prior performance, it can help them allocate students across tracks, allowing for a better match between student ability level and track level which benefits the student, both by allowing the student to learn more and by limiting the effect of PB on student performance. We show that tracking in general does not have a direct relation with performance. On the other hand, interactions between tracking and whether schools consider prior performance reveal that students in highly differentiated systems perform best when schools always take into account

prior performance to decide on student acceptance. In systems with a low number of tracks, whether schools consider prior performance has less of an impact.

The association between PB, tracking and student performance shows that equality of opportunity is best provided for in a system with a high number of tracks combined with schools always consider prior performance on accepting the student to the school. It turns out that for high-PB students in these systems, tracking weakens the positive relation between PB and performance, whereas for low PB students the (for them negative) relation between tracking and performance is lowered primarily when they attend schools that always consider prior performance. Thus it seems that high-PB students might be harmed by tracking when schools consider prior performance. The result, that tracking does not increase inequality of opportunity, if track placement is based on prior achievement, has recently been confirmed by Esser and Relikowski (2015). The authors compare student outcomes in two German states, one with strict assignment rules for track placement and one with less strict rules, and since they are able to control for prior achievement directly, they exclude selection into schools.

We argue that it is not straightforward to determine whether tracking in itself has a positive or negative effect on performance. When education system characteristics are studied, it should be taken into account that schools can have large influence on the implementation of these system characteristics and thus heterogeneous effects across schools can arise. We show that when tracking is combined with whether schools consider prior performance in accepting the student, tracking benefits both student performance and educational opportunities.

When more data become available in the future, these findings could be replicated controlling for individual prior performance (as Esser and Relikowski 2015) and by including more elaborate and specific information on the selection mechanism with which students are selected into tracks.

## Acknowledgements

We would like to thank Lex Borghans, Rolf van der Velden, Hans Heijke, Oliver Marie and Andreas Ammermueller and

the seminar participants of the CPB Netherlands Bureau for Economic Policy Analysis, the Journées Louis-André Gérard-Varet 2012, ESPE 2012, the Maastricht University economics of education lunch and the NWO-PROO group for useful suggestions.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Ammermueller, A. 2005. *Educational Opportunities and the Role of Institutions*. Discussion Paper 05-44. Mannheim: ZEW.
- Ariga, K., and G. Brunello. 2007. Does Secondary School Tracking Affect Performance? Evidence from IALS. IZA Discussion Paper 2643. Bonn: IZA.
- Brunello, G., and D. Checchi. 2007. "Does School Tracking Affect Equality of Opportunity? New International Evidence." *Economic Policy* 22: 782–861. doi:10.1111/j.1468-0327.2007.00189.x.
- Dollmann, J. 2011. "Verbindliche und unverbindliche Grundschulempfehlungen und soziale Ungleichheiten am ersten Bildungsübergang." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 63: 595–621. doi:10.1007/s11577-011-0148-z.
- Dronkers, J., R. Van Der Velden, and A. Dunne. 2012. "Why are Migrant Students Better off in Certain Types of Educational Systems or Schools than in Others?" *European Educational Research Journal* 11: 11–44. doi:10.2304/eeerj.
- Dustmann, C. 2004. "Parental Background, Secondary School Track Choice, and Wages." *Oxford Economic Papers* 56: 209–230. 10.1093/oep/gpf048
- Elk, R. V., M. van Der Steeg, and D. Webbink. 2011. "Does the Timing of Tracking Affect Higher Education Completion?" *Economics of Education Review* 30: 1009–1021. doi:10.1016/j.econedurev.2011.04.014.
- Esser, H., and I. Relikowski. 2015. "Is Ability Tracking (Really) Responsible for Educational Inequalities in Achievement? A Comparison between the Country States Bavaria and Hesse in Germany." *IZA Discussion Paper no. 9082*, Bonn: Iza.
- Fuchs, T., and L. Woessmann. 2007. "What Accounts for International Differences in Student Performance? A Re-Examination Using PISA Data." *Empirical Economics* 32: 433–464. doi:10.1007/s00181-006-0087-0.
- Hanushek, E., and L. Woessmann. 2006. "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries." *Economic Journal* 116: C63–C76. doi:10.1111/j.1468-0297.2006.01076.x.
- Lucas, S. 1999. *Tracking Inequality: Stratification and Mobility in American High Schools*, *Sociology of Education Series*. New York: Teachers College Press.
- OECD. 2007. *Science Competencies for Tomorrow's World*. Paris: OECD.
- OECD. 2010. *PISA 2009 at a Glance*. Paris: OECD.
- Sacerdote, B. 2011. "Peer Effects in Education: How Might They Work How Big Are They and How Much Do We Know Thus Far." In *Handbook of the Economics of Education*, edited by E. Hanushek, S. Machin, and L. Woessmann, Vol. 3, 249–277. Elsevier. doi:10.1016/B978-0-444-53429-3.00004-1.
- Sacerdote, B. 2014. "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?" *Annual Review of Economics* 6: 253–272. doi:10.1146/annurev-economics-071813-104217.
- Schuetz, G., H. Ursprung, and L. Woessmann. 2008. "Education Policy and Equality of Opportunity." *Kyklos* 61: 279–308. doi:10.1111/j.1467-6435.2008.00402.x.
- Walldinger, F. 2006. "Does tracking affect the importance of family background on students' test scores?" Unpublished paper, *mimeo*, London: CEPR.
- World Bank. 2012. *World Development Indicators and Global Development Finance Database*. Washington, DC: World Bank.